Automated prediction of cardiorespiratory deterioration in patients with single-ventricle parallel circulation: A multicenter validation study

Craig G. Rusin, PhD, a Sebastian I. Acosta, PhD, a Kenneth M. Brady, MD, b Eric Vu, MD, b Carly Scahill, MD, c Brian Fonseca, MD, c Cindy Barrett, MD, c Janet Simsic, MD, d Andrew R. Yates, MD, d Brenna Klepczynski, RN, BSN, e William J. Gaynor, MD, e and Daniel J. Penny, MD, PhD, MHA a

ABSTRACT

Objectives: Patients with single-ventricle physiology have a significant risk of cardiorespiratory deterioration between their first- and second-stage palliation surgeries. Detection of deterioration episodes may allow for early intervention and improved outcomes.

Methods: A prospective study was executed at Nationwide Children’s Hospital, Children’s Hospital of Philadelphia, and Children’s Hospital Colorado to collect physiologic data of subjects with single ventricle physiology during all hospitalizations between neonatal palliation and II surgeries using the Sickbay software platform (Medical Informatics Corp). Timing of cardiorespiratory deterioration events was captured via chart review. The predictive algorithm previously developed and validated at Texas Children’s Hospital was applied to these data without retraining. Standard metrics such as receiver operating curve area, positive and negative likelihood ratio, and alert rates were calculated to establish clinical performance of the predictive algorithm.

Results: Our cohort consisted of 58 subjects admitted to the cardiac intensive care unit and stepdown units of participating centers over 14 months. Approximately 28,991 hours of high-resolution physiologic waveform and vital sign data were collected using the Sickbay. A total of 30 cardiorespiratory deterioration events were observed. The risk index metric generated by our algorithm was found to be both sensitive and specific for detecting impending events one to two hours in advance of overt extremis (receiver operating curve = 0.927).

Conclusions: Our algorithm can provide a 1- to 2-hour advanced warning for 53.6% of all cardiorespiratory deterioration events in children with single-ventricle physiology during their initial postop course as well as interstage hospitalizations after stage I palliation. (JTCVS Open 2023; -:1-6)

Patients with single-ventricle lesions account for 2% to 3% of all congenital heart disease a but are responsible for up to 25% to 40% of all neonatal cardiac deaths. b-d The time period of greatest risk for these patients is in the immediate postoperative period after neonatal palliation. e,f During this time period, patients have a parallel circulatory system in which a single functioning ventricle perfuses both the systemic and pulmonary circuits.

From the aDepartment of Pediatrics—Cardiology, Baylor College of Medicine, Texas Children’s Hospital, Houston, Tex; bDepartment of Anesthesiology, Northwestern University, Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, Ill; cDepartment of Pediatrics—Cardiology, Children’s Hospital Colorado, Aurora, Colo; dDepartment of Pediatrics—Cardiology, Nationwide Children’s Hospital, Columbus, Ohio; and eDepartment of Cardiovascular Surgery, Children’s Hospital of Philadelphia, Philadelphia, Pa.

Received for publication Jan 30, 2023; revisions received April 13, 2023; accepted for publication May 2, 2023.

Address for reprints: Craig G. Rusin, PhD, 1102 Bates St, Feigin Building Suite 430.03, Houston, TX 77030 (E-mail: cgrusin@bcm.edu).

2666-2736

Copyright © 2023 The Authors. Published by Elsevier Inc. on behalf of The American Association for Thoracic Surgery. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.1016/j.xjon.2023.05.012
Abbreviations and Acronyms
ECG = electrocardiogram
HR = heart rate
HRV = heart rate variability
PVC = frequency of premature ventricular contractions
PVI = pleth variability index
RI = risk index
ST = ST-segment elevation
STV = ST-segment variability
TPR = true positive rate

Such a circuit is inherently unstable, as changes in pulmonary or systemic vascular resistances can affect systemic oxygen delivery significantly, resulting sudden and unexpected patient deterioration. As a result, mortality between neonatal palliation and stage II surgeries for this population is approximately 15%, and studies suggest that cardiorespiratory arrests account for 63% of the identifiable causes of death during this time period.7

We have previously developed and tested an automated algorithm that can predict impending clinical deterioration events in this population, during their hospitalization, 1 to 2 hours in advance of overt symptoms with high sensitivity and specificity in a large single-center study.8,9 In this report, we present the results of a multicenter validation study designed to measure the performance of the same algorithm when exposed to different clinical practice patterns at several tertiary pediatric heart centers in the United States. To our knowledge, this is the first fully automated algorithm for predicting acute cardiorespiratory deterioration events not only continuously and in real time but also validated in an independent, multicenter study. Although generalized risk models have been developed, none currently provide a warning of the immediate risk of an acute event.10-12

METHODS
Subjects, Deterioration Events, and Data Collection
The study was designed as a prospective, observational, multicenter validation study. Study subjects were recruited, consented, and prospectively enrolled at 3 leading pediatric heart centers in the United States between May 9, 2018, and September 24, 2019. Participating centers included Nationwide Children’s Hospital, Children’s Hospital of Philadelphia, and Children’s Hospital Colorado. The institutional review board or equivalent ethics committee of the authors’ institutions approved the study protocol and publication of data (institutional review board #H-33066, approval date July 3, 2013). The patient(s) provided informed written consent for the publication of the study data. The inclusion criteria for the study were infants born with morphologic hypoplasia of either the right or left ventricles of the heart and underwent neonatal palliation surgery to establish parallel circulation. Study subjects were admitted to the cardiovascular intensive care units at their respective centers immediately after their neonatal surgical palliation. Study enrollment terminated either at the time of stage II palliation surgery, patient discharge home, patient withdrawal, or death.

The physiological data from all study subjects were recorded from their postoperative Norwood hospitalization and any hospitalization during interstage period in both the intensive care units and step-down units. If a subject was discharged and later readmitted, the physiologic data from this readmission was included in the study (ie, the patient had not received stage II palliation). Physiologic data were acquired continuously using the Food and Drug Administration–cleared, Sickbay patient monitoring and analytics platform (510k-K143304; Medical Informatics Corporation). Physiologic signals included high-resolution waveforms (ECG, photoplethysmogram) and low-resolution vital signs calculated by the bedside monitors (heart rate [HR], arterial oxygen saturation, ST-segment elevations, premature ventricular contractions). The Sickbay platform collected, processed, stored, and deidentified the data as well as executed data analysis using its research application programming interfaces and development environment.

Cardiorespiratory deterioration events were defined as either a cardiac arrest requiring cardiopulmonary resuscitation, extracorporeal membrane oxygenation, or an unplanned intubation. The approximate time of each deterioration event was first obtained via chart review and confirmed to the exact minute by clinical investigators using the physiological recordings of heart and respiratory rates. The physiological data generated by the cohort were labeled as either predeterioration data or nondeterioration data. The predeterioration data included signals recorded between 1 and 2 hours before the event. The nondeterioration data included all physiologic signals recorded more than 2 hours before or more than 48 hours after a deterioration event. Data labels were used to assess the performance of the algorithm.

Data Processing
The classification algorithm developed by Rusin and colleagues8 uses 7 input variables calculated from the physiologic signals. The 7 input metrics include HR, heart rate variability (HRV), peripheral arterial oxygen saturation, ST-segment elevation (ST), ST-segment variability (STV), frequency of premature ventricular contractions (PVC), and the pleth variability index (PVI). The HR, peripheral arterial oxygen saturation, ST, and PVC were acquired directly from the standard bedside monitor by averaging the raw signals over 5-minute windows. The HRV, STV, and PVI require additional processing steps. The HRV was processed from the ECG waveform by calculating the time lapse between R-peaks to obtain the instantaneous frequency of cardiac beats. The HRV was then defined as the standard deviation of the beat-to-beat frequencies collected over a window of 5 minutes. The STV, defined in detail in Vu and colleagues,13 quantifies the change of the ST-segment vector every 30 seconds using 3-dimensional orthogonal projections estimated from the ST-segment elevations from the V1, aVL, and II ECG leads. The magnitude of the 30-second changes is averaged over 5-minute windows. The PVI is an estimate of the amplitude modulation of the plethysmograph waveform at a frequency band around the respiration rate. As a surrogate for the pulse pressure variability, the PVI is associated with volume responsiveness.14 Using this approach, PVI was computed every minute and then averaged over 5-minute windows.

The classification algorithms developed by Rusin and colleagues8 is based on a logistic regression to compute a risk index of the form

\[
RI = \exp \left( \sum_{j=1}^{7} \alpha_j x_j - \mu_j \right) / \sigma_j
\]

where RI is the risk index associated with the input variables \(x_1, x_2, \ldots, x_7\) in the predeterioration dataset, and \(\mu_j\) and \(\sigma_j\) are used to normalize the variables. Each of the input variables \(x_j\) at a given time \(t \) were computed as deviation of the physiologic metrics from their moving baselines values as follows:

\[
x_j(t) = y_j(t) - \bar{y}_j(t)
\]
where \( y_1 = \text{HR}, \ y_2 = \text{HRV}, \ y_3 = \text{SpO}_2, \ y_4 = \text{ST}, \ y_5 = \text{STV}, \ y_6 = \text{PVC}, \) and \( y_7 = \text{PVI}, \) and \( y_{base}^j(t) \) is the moving baseline value of the metric \( j \) calculated as the average of \( y_j \) over a window of time that starts 24 hours before time \( t \) and ends 12 hours before time \( t. \) The same coefficients from Rusin and colleagues\(^8\) were used to validate this classification model with the independent data from study sites.

### Algorithm Performance

Once the logistic input variables \( (x_1, x_2, \ldots, x_7) \) were computed, the risk index \( RI \) was calculated according to equation (1) for every 5-minute interval for each subject in the cohort. The performance of this classification was quantified using the area under the receiver operating characteristic curve as a threshold-independent metric. For thresholds ranging from 0.1 to 100 to classify \( RI \) as either predeterioration or nondeterioration, we also computed the true-positive rate (TPR), false-positive rate, Matthews correlation coefficient, positive likelihood ratio, and negative likelihood ratio.

### RESULTS

The study cohort included 58 (Children’s Hospital of Philadelphia \( N = 27 \), Children’s Hospital Colorado \( N = 10 \), Nationwide Children’s Hospital \( N = 21 \)) subjects who produced usable data for the validation analysis. Study subjects experienced 30 deterioration events over the course of their hospitalization between neonatal palliation and stage II surgeries, while physiologic data were being recorded, 12 of which were found to be cardiac arrests and 18 were respiratory events. Two subjects withdrew from the study, neither of whom experienced a deterioration event. A review of discharge dates after the neonatal palliations revealed that 60\% of the events (18 of 30) occurred before the discharge date of the first hospitalization. The rest of the events occurred during rehospitalizations but before the stage II palliation. Twenty-one subjects experienced at least 1 event. Ten patients died during the study, yielding a mortality of 17.2\%. Figure 1 displays a histogram of deterioration events relative to neonatal palliation date. It was observed that 50\% of the deterioration events occurred within the first 30 days after neonatal palliation. Table 1 describes the types of neonatal palliative procedures for the cohort and the prevalence of deterioration events. The Norwood procedure was found to be the most prevalent surgery and accounted for 66\% of the neonatal palliations in the cohort.

Between neonatal palliation and stage II palliations was a median of 141 days (interquartile range, 126-155 days). For the entire cohort, the cumulative time in between neonatal palliation and stage II palliations was 206,088 hours. The cumulative time between admission and discharge dates of the first hospitalization was 59,760 hours. There were 28,991 hours with recorded physiological data.

The classification algorithm developed in Rusin and colleagues\(^8\) was evaluated on the 7 physiologic metrics from the multisite validation cohort. Since these data were labeled as either nondeterioration or predeterioration according to its proximity to deterioration events, the performance of the classification algorithm was assessed.
for cardiorespiratory deterioration events as a function of a specific alert threshold. Results indicate that 46% to 53% of all cardiorespiratory deterioration events can be predicted 1 to 2 hours in advance by the algorithm. Results indicate that cardiac events are more accurately predicted compared to respiratory events.

Table 3 illustrates the performance of the predictive model separated by study site. Although the results suggest that there is a slight increase in predictive performance in site 1 compared with the others, the model still was found to have still have significant performance, indicating that the model is generalizing well across different sites consistently.

**DISCUSSION**

Computer algorithms will never replace trained medical personnel. However, they can be used to augment and enhance the capabilities of the clinician and make the medical care environment safer for patients and more efficient for providers. This is part of an ongoing shift in critical care management referred to as virtual critical care, the idea that several layers of clinical assets (local providers, remote providers, and algorithmic surveillance) can coordinate together virtually to form a unified system of patient oversight. Algorithmic surveillance plays a key role in virtual care systems, as it provides a scalable means of identifying key moments in time when human intervention is most needed and resources can be efficiently deployed.\(^{15}\)

The ability to detect sudden acute deterioration events immediately before they occur allows the care team to take time-critical corrective actions to prevent these subclinical events from progressing to life-threatening ones. As a result, this work represents a shift in the way these patients are monitored in hospital: moving away from standard physiologic data acquisition at the bedside and

---

**FIGURE 2.** Performance characteristics. Top left, Receiver operating characteristic curve. Top right, True-positive rate (TPR) and false-positive rate (FPR) as functions of the risk index (RI) threshold. Bottom left, Matthews correlation coefficient (MCC) as a function of the RI threshold. Bottom right, Positive likelihood ratio (PLR) and negative likelihood ratio (NLR) as a function of the RI threshold.
instead focusing on automating physiologic data interpretation. In this way, automated monitoring can provide a scalable layer of advanced algorithmic surveillance for patients. Although no algorithm will ever replace a physician or nurse, algorithms working in real-time with the care team can make patient surveillance more efficient and effective than either can achieve on their own.

Our study demonstrated that the majority of cardiorespiratory deterioration events that occurred during the hospitalization of single-ventricle population before stage II palliation can be predicted 1 to 2 hours in advance of overt symptoms using automated, algorithmic surveillance. With an update frequency of 5 minutes, the algorithm can detect subtle, moment-to-moment physiologic changes that may occur throughout the hospitalization, such as a bath. Furthermore, our study demonstrated that the physiologic signs of predeterioration in this population are not specific to a particular clinical environments or practice patterns. The algorithm appears to be translatable across institutions and clinical environments as both intensive care units and step-down units were included in the analysis. Cardiac events seem to be more consistently predictable than respiratory deteriorations. This is likely the case as cardiac events are often more discrete and unambiguous compared to respiratory events, which may have more clinical subjectivity around their recognition. The natural imprecision in the recognition of respiratory events leads to lower predictive performance.

There is, however, a tradeoff between predictive performance and the number of alerts that are generated in the care environment. The greater the alerting threshold, the lower the number of alerts generated, but the lower the TPR. The selection of the alerting threshold allows an institution to balance the predictive performance of the algorithm with the load that the algorithm will place on the care team. For example, at an alert threshold of 1, the median number of alerts generated by the algorithm is 2.5 alerts per patient per day, but the TPR is 54%. What this means is that if the care team responded to these 2.5 alerts per patient per day, they would be able to anticipate 54% of all deterioration events 1 to 2 hours before they happen. In context, there are typically 100 to 400 alarms per patient per day in this population generated from existing bedside devices.

Such an algorithm may have utility even when it fails to forecast an event. Fundamentally, this algorithm has been trained to recognize the physiological fingerprint associated with the subclinical precursors of a critical deterioration event. If a patient experiences an event, and no such clinical precursors are present (as would be indicated by a low value of the risk index), then this may suggest a different etiology, such as shunt thrombosis.

There are several limitations to this study. First, the predictive algorithm does not explicitly consider the types of cardiac lesion, post-surgical anatomy, laboratory values, co-morbidities, and other health characteristics. Although all the patients included in the study have parallel circulation, there are differences in physiology between anatomies that may influence the patient’s susceptibility to cardiopulmonary deteriorations and response to interventions.16-18 The incorporation of these categorical features into a

### Table 2. Performance characteristics for the evaluation of the predictive model

<table>
<thead>
<tr>
<th>Risk index threshold</th>
<th>Matthews correlation coefficient</th>
<th>TPR overall, %</th>
<th>TPR cardiac events, %</th>
<th>TPR respiratory events, %</th>
<th>FPR overall, %</th>
<th>Alarms per day per patient, median (interquartile range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.27</td>
<td>53.61</td>
<td>67.86</td>
<td>31.48</td>
<td>2.07</td>
<td>2.5 (1.5-5.4)</td>
</tr>
<tr>
<td>1.5</td>
<td>0.30</td>
<td>46.37</td>
<td>56.40</td>
<td>17.23</td>
<td>1.27</td>
<td>1.3 (0.1-3.8)</td>
</tr>
<tr>
<td>2</td>
<td>0.29</td>
<td>39.14</td>
<td>48.54</td>
<td>11.08</td>
<td>0.89</td>
<td>1.1 (0.0-2.4)</td>
</tr>
<tr>
<td>3</td>
<td>0.32</td>
<td>34.78</td>
<td>45.34</td>
<td>7.41</td>
<td>0.56</td>
<td>0.4 (0.0-1.4)</td>
</tr>
<tr>
<td>4</td>
<td>0.33</td>
<td>31.94</td>
<td>39.89</td>
<td>7.41</td>
<td>0.42</td>
<td>0.2 (0.0-1.0)</td>
</tr>
<tr>
<td>5</td>
<td>0.34</td>
<td>30.25</td>
<td>35.75</td>
<td>7.41</td>
<td>0.33</td>
<td>0.0 (0.0-0.6)</td>
</tr>
<tr>
<td>6</td>
<td>0.31</td>
<td>26.01</td>
<td>32.47</td>
<td>5.56</td>
<td>0.28</td>
<td>0.0 (0.0-0.5)</td>
</tr>
</tbody>
</table>

TPR, True-positive rate; FPR, false-positive rate.

### Table 3. Performance characteristics for the evaluation of the predictive model broken down by study site

<table>
<thead>
<tr>
<th>Risk index threshold</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPR, %</td>
<td>FPR, %</td>
<td>TPR, %</td>
</tr>
<tr>
<td>1</td>
<td>65.6</td>
<td>3.4</td>
<td>51.5</td>
</tr>
<tr>
<td>1.5</td>
<td>65.6</td>
<td>2.8</td>
<td>39.1</td>
</tr>
<tr>
<td>2</td>
<td>62.4</td>
<td>1.8</td>
<td>32.6</td>
</tr>
<tr>
<td>3</td>
<td>56.3</td>
<td>1.2</td>
<td>23.8</td>
</tr>
</tbody>
</table>

TPR, True-positive rate; FPR, false-positive rate.
predictive algorithm may require a much larger sample size not presently available, and the algorithm was found to generalize well across centers and types of cardiac lesions. Second, pulmonary deterioration and the precise time for intubation may be ambiguous as they depend to some extent on clinical judgement. Therefore, any algorithm attempting to predict these events with high precision may have difficulties. Finally, although the algorithm may be able to predict deterioration events in the near future accurately, the clinical benefit of this prediction has yet to be determined. Measuring the potential outcomes benefit related to this algorithm paired with an intervention protocol is planned as a future multicenter clinical trial.

CONCLUSIONS
We have demonstrated that there exist subtle, yet detectable, physiologic changes in the hours preceding cardiopulmonary deterioration in children with single-ventricle physiology. We have validated an algorithm that was trained to detect these changes in this population. A multicenter trial showed that this algorithm was able to predict 53.6% of all cardiorespiratory deterioration 1 to 2 hours before they happened with only 2.5 alarms being generated per patient per day.

Conflict of Interest Statement
Dr Rusin is a co-founder of Medical Informatics Corp. All other authors report no conflicts of interest.

The Journal policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

References

Key Words: clinical deterioration, prediction algorithm, single-ventricle physiology, arrest prediction, data mining, forecasting